The Use of Personal Income Tax Data Aggregated at Grid Level in the Dwelling Price Index Calculation of Slovenia

Branko Pavlin, PhD
Senior Adviser, Statistical Office of the Republic of Slovenia

Abstract

Dwelling prices are influenced by a number of factors and each of them has a different significance or impact.

The Statistical Office of the Republic of Slovenia (SORS) has recently implemented the hedonic method for calculating dwelling price indices. In this approach, various dwelling characteristics are used as »explanation variables« in the multiple regression analyses. The output regression coefficient of a single dwelling characteristic indicates the proportional influence of that characteristic to the dwelling price.

Besides physical characteristics (i.e. living area, age, etc.), location characteristics have a considerable impact on real estate price formation. One of them - »Wealth of the neighbourhood« - based on average gross personal income tax data tends to explain the impact of the wealth of neighbourhood's residents on dwelling price levels. Neighbourhoods may be spatially defined and sized differently: as small size administrative areas or as regular grid cells of different side size. We experimented with both ways and examined the impact on the dwelling price indices.

1. Introduction: price index calculation supported by grid data

The hedonic approach is basically a two-stage procedure. In the first stage the hedonic regression analysis is applied to establish the importance of individual characteristics of the dwelling on its price. Dwelling characteristics values are transformed into real or dummy »explanation variables« and used in the multiple regression analyses. The outcome regression coefficients (»r«) indicate the implicit price of dwelling characteristics. In the second stage these regression coefficients are used in several known approaches (formulas) for calculating the index.

In this study main attention is given to one of the location explanation variables named "wealth of the neighbourhood". We experimented with this variable by spatially shaping it in three different ways: as census districts, as 1x1km grid or as 0.5x0.5km grid. In successive regression analyses different territorial units of this variable were included while all other variables based on dwelling characteristics remained strictly unchanged during the experiment. Until recently, census districts were normally used as analytical territorial units in SORS's price index calculations. This experiment aimed to discover whether the use of grid data could improve the regression analyses reliability.

1

2. The explanation variable »Wealth of the neighbourhood«

Our thesis was that in neighbourhoods where people are generally wealthier dwellings are more expensive and vice versa. Wealth of the neighbourhood is therefore seen as a synthetic indicator for dwelling price formation, which comprises construction-technical characteristics of buildings in the neighbourhood, visual qualities, availability of services, transport situation and social reputation of the neighbourhood. Therefore, dwelling prices variability can to a certain extent be explained by the data on the wealth of residents of the neighbourhood.
»Wealth of the neighbourhood« in this study was defined as the average 2005-2006 gross personal income tax, calculated per person, for residents of the neighbourhood liable to pay tax.

3. Spatial shaping of the variable »Wealth of the neighbourhood«

We applied the concept of »wealth of the neighbourhood« to three differently shaped and sized spatial analytical units: census districts, 1x1km grid cells and 0.5x0.5km grid cells. There are 17,026 census districts in Slovenia and they are shaped in such a way that each one comprises a maximum of 250 inhabitants. In densely populated areas census districts are numerous and small, while in sparsely populated areas census districts may be very large.
There are approximately 20,000 1x1km grid cells in Slovenia and 80,000 0.5x0.5km grid cells.
It is clearly seen from Table1 (below) that the selection of spatial units has an impact on values of the explanation variable "wealth of the neighbourhood".

Table1: Average values of the explanation variable "wealth of the neighbourhood" calculated at three levels of territorial breakdown and three types of spatial analytical areas (applied only in areas with transactions of dwellings in 2007)

		Census district €	Grid 1x1km €	Grid 0.5x0.5km €
Ljubljana municipality - dwellings	Average value (€)	13022	13075	11830
	Standard deviation (€)	2626	1490	1857
Rest of the country - dwellings	Average value (€)	10783	11009	10962
	Standard deviation (€)	1638	1116	1302
Slovenia - houses	Average value (€)	11365	11522	11517
	Standard deviation (€)	2021	1687	2649

Not only average values but also the spatial distribution of variable values differ within three types of spatial analytical units (Maps 1 to 3 in Appendix).

4. The Case study: Dwelling price indices of Ljubljana, Rest of the country and Slovenia as a whole

Firstly, by using analytical methods of geographic information systems (GIS), data on every single dwelling transaction were supplemented by the information on »wealth of the neighbourhood«. After that these data were employed in the hedonic regression analyses besides other physical and location characteristics data of dwellings.

The regression analyses and the calculation of price indices have been performed at three levels of territorial breakdown: the urban municipality of Ljubljana (1489 apartments), Rest of the country (4751 apartments) and Slovenia as whole (2165 family houses). At each territorial level the explanation variable "Wealth of the neighbourhood" has been calculated by census districts, 1km grid cells and 0.5km grid cells. A total of 7 explanation variables have been used in the regression analyses of apartment prices in Ljubljana, 20 variables in Rest of the country and 14 variables for family house prices at the territory of Slovenia.

 Results of the comparative hedonic regression analyses are presented in Table2.

Table2: Dwelling price indices calculated by using different spatial shaping of the explanation variable »wealth of the neighbourhood« at three territorial breakdown levels: Ljubljana, Rest of the country and Slovenia as whole

		Price indices of dwellings / spatial units		
		Census district	Grid 1x1km	Grid 0.5x0.5km
APPARTMENTS	q1-2007	100,0	100,0	100,0
Ljubljana municipality	q2-2007	103,5	103,7	103,6
	q3-2007	105,4	105,4	105,8
	q42007	105,9	106,5	106,5
R^2- Multiple regression coefficients [1]		**0,815**	**0,813**	**0,813**
R- Partial regression coefficients (in %) [2]		2,96	1,69	2,02
APPARTMENTS	q1-2007	100,0	100,0	100,0
Rest of the country	q2-2007	104,8	104,6	104,3
	q3-2007	101,8	101,7	101,4
	q42007	108,5	108,5	108,3
R^2- Multiple regression coefficients		**0,534**	**0,533**	**0,537**
R- Partial regression coefficients (in %)		4,28	4,47	5,01
FAMILY HOUSES	q1-2007	100,0	100,0	100,0
Slovenia	q2-2007	98,2	98,0	98,1
	q3-2007	109,1	107,0	107,0
	q42007	107,8	107,3	107,6
R^2- Multiple regression coefficients		**0,656**	**0,658**	**0,642**
R- Partial regression coefficients (in %)		9,90	33,70	6,67

[1] The **multiple regression coefficients – R^2** indicate the proportion of dwelling prices variance that is explained by all explanation variables. The maximum value of 1 is in practice

never reached and would mean that 100% of price variability is explained by explanation variables. Higher figures of R^2 indicate more successful regression.

^{2} **Partial R** tells us what portion of price variability can be explained by a single explanation variable.

Findings:
1. Spatial shaping of the analytical units of the variable »Wealth of the neighbourhood« does have some impact on hedonic regression coefficients and thus in the final consequence on price indices! This impact is rather small or moderate, and is analogue to the overall importance of the explanation variable »Wealth of the neighbourhood« in the regression analyses, measured by the partial regression coefficients.

2. The **comparison of price indices** calculated by the use of census districts as spatial-analytical units and those calculated by the use of grid-shaped areas has shown the following:
   - Price indices of dwellings in Ljubljana are 0.1% to 0.6% higher when grid-shaped areas are used in regression analyses and 0.1% to 0.5% lower when calculated for Rest of the country
   - Price indices of family houses are lower (from 0.1% to 1.9%) when grid-shaped variable »Wealth...) is used.

3. Slightly lower (-0.2) **multiple regression coefficients** are reached when 1x1km grid and 0.5x0.5km grid-shaped variable »Wealth of the neighbourhood« is used in regression analyses of apartments in Ljubljana. This means that census districts are appropriate spatial analytical units for the regression analyses in Ljubljana and probably also for urban areas in general!

4. In the hedonic regression analyses of apartment prices in Rest of the country the multiple regression coefficients are slightly higher for 0.5x0.5km grid in comparison with census districts and 1x1km grid. This means that 0.5x0.5km grid could be appropriate for regression analyses of non-urban areas. Grid 1x1km seems to be too coarse for apartment price analyses either in urban or non-urban areas!

5. However, in the regression analyses of family houses for the territory of whole Slovenia the highest multiple regression coefficient has been reached with 1x1km grided variable »Wealth of the neighbourhood«.

Conclusions:

Inclusion of explanation variable "Wealth of the neighbourhood" calculated by 1km or 0.5km grid cells instead of by census districts has not increased significantly the overall accuracy of hedonic regression analyses, measured indirectly by the multiple regression coefficients. However, small improvement has been achieved by the use of 0.5x0.5km grid cells as analytical units in the regression of apartment prices outside urban area of Ljubljana and by the use of 1x1km grid for family house price analyses at the territorial level of the whole country. For urban areas census districts

seem to be the most appropriate spatial-analytical units for the variable »Wealth of the neighbourhood«. To draw up any other conclusion longer time series should be observed and analyzed at least.

This small and non-representative experiment has opened a way for further researching. It would be interesting to discover if spatial aggregating of 0.5x0.5km grid data into more dwelling price-homogeneous analytical areas would enhance the hedonic regression analyses, especially in the non-urban areas.

Appendix: Maps